In [2]:

```
#Improts

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import sqlite3
import seaborn as sns
import matplotlib.pyplot as plt

# Input data files are available in the "../input/" directory.
# For example, running this (by clicking run or pressing Shift+Enter) will list the files
in the input directory
```

In [3]:

```
conn = sqlite3.connect('database.sqlite')

tables = pd.read_sql("""SELECT *
                        FROM sqlite_master
                        WHERE type='table';""", conn)
tables
```

Out[3]:

	type	name	tbl_name	rootpage	sql
0	table	sqlite_sequence	sqlite_sequence	4	CREATE TABLE sqlite_sequence(name,seq)
1	table	Player_Attributes	Player_Attributes	11	CREATE TABLE "Player_Attributes" (\n\t`id`\tIN...
2	table	Player	Player	14	CREATE TABLE `Player` (\n\t`id`\tINTEGER PRIMA...
3	table	Match	Match	18	CREATE TABLE `Match` (\n\t`id`\tINTEGER PRIMAR...
4	table	League	League	24	CREATE TABLE `League` (\n\t`id`\tINTEGER PRIMA...
5	table	Country	Country	26	CREATE TABLE `Country` (\n\t`id`\tINTEGER PRIM...
6	table	Team	Team	29	CREATE TABLE "Team" (\n\t`id`\tINTEGER PRIMARY...
7	table	Team_Attributes	Team_Attributes	2	CREATE TABLE `Team_Attributes` (\n\t`id`\tINTE...

In [4]:

```
countries = pd.read_sql("""SELECT *
                        FROM Country;""", conn)
countries
```

Out[4]:

	id	name
0	1	Belgium
1	1729	England
2	4769	France
3	7809	Germany
4	10257	Italy
5	13274	Netherlands
6	15722	Poland
7	17642	Portugal
8	19694	Scotland
9	21518	Spain
10	24558	Switzerland

In [5]:

```
leagues = pd.read_sql("""SELECT *
                         FROM League
                         JOIN Country ON Country.id = League.country_id;""", conn)
leagues
```

Out[5]:

	id	country_id	name	id	name
0	1	1	Belgium Jupiler League	1	Belgium
1	1729	1729	England Premier League	1729	England
2	4769	4769	France Ligue 1	4769	France
3	7809	7809	Germany 1. Bundesliga	7809	Germany
4	10257	10257	Italy Serie A	10257	Italy
5	13274	13274	Netherlands Eredivisie	13274	Netherlands
6	15722	15722	Poland Ekstraklasa	15722	Poland
7	17642	17642	Portugal Liga ZON Sagres	17642	Portugal
8	19694	19694	Scotland Premier League	19694	Scotland
9	21518	21518	Spain LIGA BBVA	21518	Spain
10	24558	24558	Switzerland Super League	24558	Switzerland

In [6]:

```
teams = pd.read_sql("""SELECT *
                       FROM Team
                       ORDER BY team_long_name
                       LIMIT 10;""", conn)
teams
```

Out[6]:

	id	team_api_id	team_fifa_api_id	team_long_name	team_short_name
0	16848	8350	29	1. FC Kaiserslautern	KAI
1	15624	8722	31	1. FC Köln	FCK
2	16239	8165	171	1. FC Nürnberg	NUR
3	16243	9905	169	1. FSV Mainz 05	MAI
4	11817	8576	614	AC Ajaccio	AJA
5	11074	108893	111989	AC Arles-Avignon	ARL
6	49116	6493	1714	AC Bellinzona	BEL
7	26560	10217	650	ADO Den Haag	HAA
8	9537	8583	57	AJ Auxerre	AUX
9	9547	9829	69	AS Monaco	MON

In [7]:

```
detailed_matches = pd.read_sql("""SELECT Match.id,
                                         Country.name AS country_name,
                                         League.name AS league_name,
                                         season,
                                         stage,
                                         date,
                                         HT.team_long_name AS  home_team,
                                         AT.team_long_name AS away_team,
                                         home_team_goal,
                                         away_team_goal
```

```
                                FROM Match
                                JOIN Country on Country.id = Match.country_id
                                JOIN League on League.id = Match.league_id
                                LEFT JOIN Team AS HT on HT.team_api_id = Match.home_team
_api_id
                                LEFT JOIN Team AS AT on AT.team_api_id = Match.away_team
_api_id
                                WHERE country_name = 'Spain'
                                ORDER by date
                                LIMIT 10;""", conn)
detailed_matches
```

Out[7]:

	id	country_name	league_name	season	stage	date	home_team	away_team	home_team_goal	away_team_goa
0	21518	Spain	Spain LIGA BBVA	2008/2009	1	2008-08-30 00:00:00	Valencia CF	RCD Mallorca	3	
1	21525	Spain	Spain LIGA BBVA	2008/2009	1	2008-08-30 00:00:00	RCD Espanyol	Real Valladolid	1	
2	21519	Spain	Spain LIGA BBVA	2008/2009	1	2008-08-31 00:00:00	CA Osasuna	Villarreal CF	1	
3	21520	Spain	Spain LIGA BBVA	2008/2009	1	2008-08-31 00:00:00	RC Deportivo de La Coruña	Real Madrid CF	2	
4	21521	Spain	Spain LIGA BBVA	2008/2009	1	2008-08-31 00:00:00	CD Numancia	FC Barcelona	1	
5	21522	Spain	Spain LIGA BBVA	2008/2009	1	2008-08-31 00:00:00	Racing Santander	Sevilla FC	1	
6	21523	Spain	Spain LIGA BBVA	2008/2009	1	2008-08-31 00:00:00	Real Sporting de Gijón	Getafe CF	1	
7	21524	Spain	Spain LIGA BBVA	2008/2009	1	2008-08-31 00:00:00	Real Betis Balompié	RC Recreativo	0	
8	21526	Spain	Spain LIGA BBVA	2008/2009	1	2008-08-31 00:00:00	Athletic Club de Bilbao	UD Almería	1	
9	21527	Spain	Spain LIGA BBVA	2008/2009	1	2008-08-31 00:00:00	Atlético Madrid	Málaga CF	4	

In [8]:

```
leages_by_season = pd.read_sql("""SELECT Country.name AS country_name,
                                      League.name AS league_name,
                                      season,
                                      count(distinct stage) AS number_of_stages,
                                      count(distinct HT.team_long_name) AS number_of_t
eams,
                                      avg(home_team_goal) AS avg_home_team_scors,
                                      avg(away_team_goal) AS avg_away_team_goals,
                                      avg(home_team_goal-away_team_goal) AS avg_goal_d
if,
                                      avg(home_team_goal+away_team_goal) AS avg_goals,

                                      sum(home_team_goal+away_team_goal) AS total_goal
s
                                FROM Match
```

```
                                JOIN Country on Country.id = Match.country_id
                                JOIN League on League.id = Match.league_id
                                LEFT JOIN Team AS HT on HT.team_api_id = Match.home_team
_api_id
                                LEFT JOIN Team AS AT on AT.team_api_id = Match.away_team
_api_id
                                WHERE country_name in ('Spain', 'Germany', 'France', 'It
aly', 'England')
                                GROUP BY Country.name, League.name, season
                                HAVING count(distinct stage) > 10
                                ORDER BY Country.name, League.name, season DESC
                                ;""", conn)
leages_by_season
```

Out[8]:

	country_name	league_name	season	number_of_stages	number_of_teams	avg_home_team_scors	avg_away_team_goa
0	England	England Premier League	2015/2016	38	20	1.492105	1.20789
1	England	England Premier League	2014/2015	38	20	1.473684	1.09210
2	England	England Premier League	2013/2014	38	20	1.573684	1.19473
3	England	England Premier League	2012/2013	38	20	1.557895	1.23947
4	England	England Premier League	2011/2012	38	20	1.589474	1.21578
5	England	England Premier League	2010/2011	38	20	1.623684	1.17368
6	England	England Premier League	2009/2010	38	20	1.697368	1.07368
7	England	England Premier League	2008/2009	38	20	1.400000	1.07894
8	France	France Ligue 1	2015/2016	38	20	1.436842	1.08947
9	France	France Ligue 1	2014/2015	38	20	1.410526	1.08157
10	France	France Ligue 1	2013/2014	38	20	1.415789	1.03947
11	France	France Ligue 1	2012/2013	38	20	1.468421	1.07631
12	France	France Ligue 1	2011/2012	38	20	1.473684	1.04210
13	France	France Ligue 1	2010/2011	38	20	1.342105	1.00000
14	France	France Ligue 1	2009/2010	38	20	1.389474	1.02105
15	France	France Ligue 1	2008/2009	38	20	1.286842	0.97105
16	Germany	Germany 1. Bundesliga	2015/2016	34	18	1.565359	1.26470
17	Germany	Germany 1. Bundesliga	2014/2015	34	18	1.588235	1.16666
18	Germany	Germany 1.	2013/2014	34	18	1.743266	1.41176

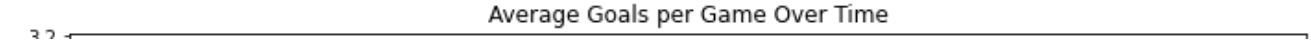
	country_name	league_name	season	number_of_stages	number_of_teams	avg_home_team_scors	avg_away_team_goa
19	Germany	Germany 1. Bundesliga	2012/2013	34	18	1.591503	1.34313
20	Germany	Germany 1. Bundesliga	2011/2012	34	18	1.660131	1.19934
21	Germany	Germany 1. Bundesliga	2010/2011	34	18	1.647059	1.27451
22	Germany	Germany 1. Bundesliga	2009/2010	34	18	1.513072	1.31699
23	Germany	Germany 1. Bundesliga	2008/2009	34	18	1.699346	1.22222
24	Italy	Italy Serie A	2015/2016	38	20	1.471053	1.10526
25	Italy	Italy Serie A	2014/2015	38	20	1.498681	1.18733
26	Italy	Italy Serie A	2013/2014	38	20	1.536842	1.18684
27	Italy	Italy Serie A	2012/2013	38	20	1.494737	1.14473
28	Italy	Italy Serie A	2011/2012	38	20	1.511173	1.07262
29	Italy	Italy Serie A	2010/2011	38	20	1.431579	1.08157
30	Italy	Italy Serie A	2009/2010	38	20	1.542105	1.06842
31	Italy	Italy Serie A	2008/2009	38	20	1.521053	1.07894
32	Spain	Spain LIGA BBVA	2015/2016	38	20	1.618421	1.12631
33	Spain	Spain LIGA BBVA	2014/2015	38	20	1.536842	1.11842
34	Spain	Spain LIGA BBVA	2013/2014	38	20	1.631579	1.11842
35	Spain	Spain LIGA BBVA	2012/2013	38	20	1.686842	1.18421
36	Spain	Spain LIGA BBVA	2011/2012	38	20	1.678947	1.08421
37	Spain	Spain LIGA BBVA	2010/2011	38	20	1.636842	1.10526
38	Spain	Spain LIGA BBVA	2009/2010	38	20	1.600000	1.11315
39	Spain	Spain LIGA BBVA	2008/2009	38	20	1.660526	1.23684

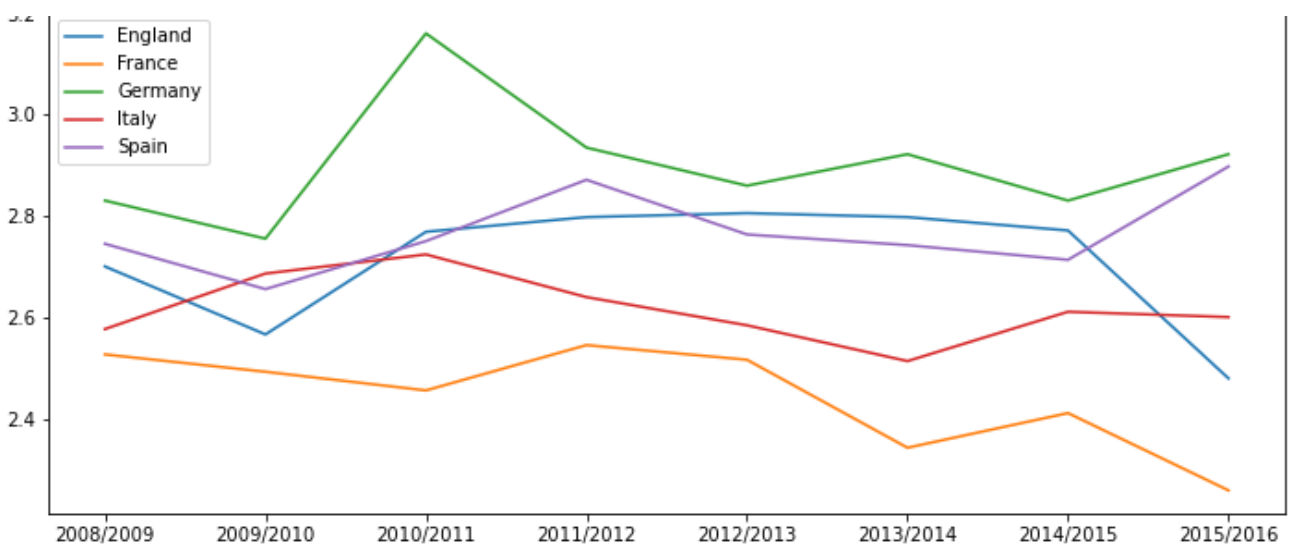
In [11]:

```
df = pd.DataFrame(index=np.sort(leages_by_season['season'].unique()), columns=leages_by_
season['country_name'].unique())

df.loc[:,'Germany'] = list(leages_by_season.loc[leages_by_season['country_name']=='German
y','avg_goals'])
df.loc[:,'Spain']   = list(leages_by_season.loc[leages_by_season['country_name']=='Spain
','avg_goals'])
df.loc[:,'France']   = list(leages_by_season.loc[leages_by_season['country_name']=='Franc
e','avg_goals'])
df.loc[:,'Italy']   = list(leages_by_season.loc[leages_by_season['country_name']=='Italy
','avg_goals'])
df.loc[:,'England']   = list(leages_by_season.loc[leages_by_season['country_name']=='Engl
and','avg_goals'])

df.plot(figsize=(12,5),title='Average Goals per Game Over Time')
```

Out[11]:

```
<AxesSubplot:title={'center':'Average Goals per Game Over Time'}>
```



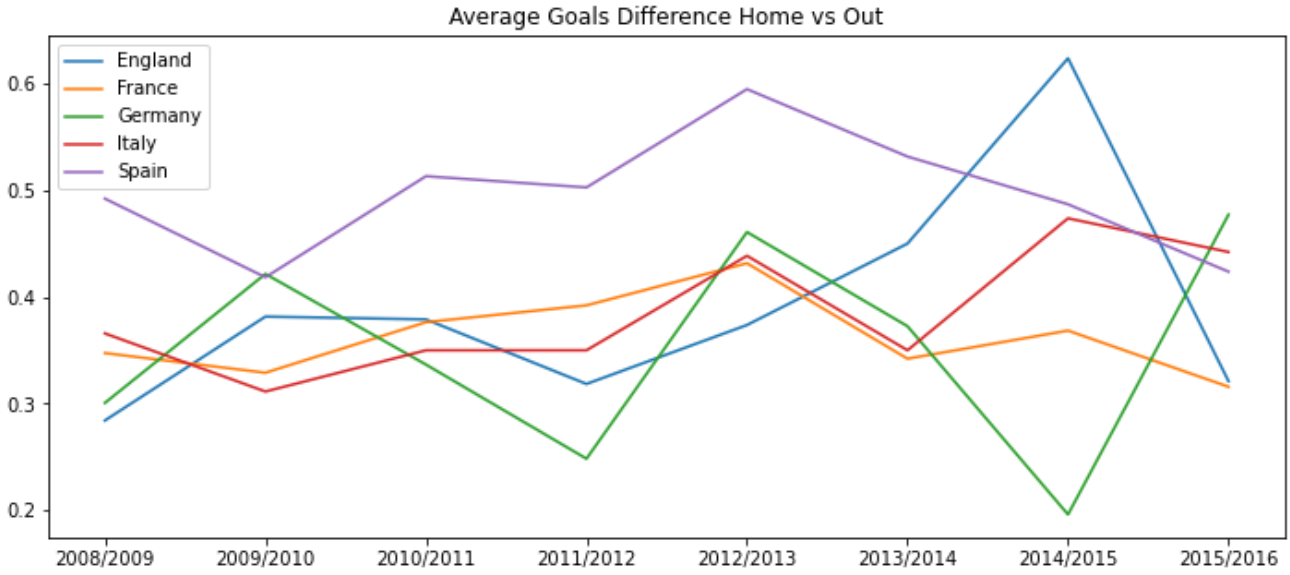


In [12]:

```
df = pd.DataFrame(index=np.sort(leages_by_season['season'].unique()), columns=leages_by_
season['country_name'].unique())

df.loc[:,'Germany'] = list(leages_by_season.loc[leages_by_season['country_name']=='German
y','avg_goal_dif'])
df.loc[:,'Spain']   = list(leages_by_season.loc[leages_by_season['country_name']=='Spain
','avg_goal_dif'])
df.loc[:,'France']   = list(leages_by_season.loc[leages_by_season['country_name']=='Franc
e','avg_goal_dif'])
df.loc[:,'Italy']   = list(leages_by_season.loc[leages_by_season['country_name']=='Italy
','avg_goal_dif'])
df.loc[:,'England']   = list(leages_by_season.loc[leages_by_season['country_name']=='Engl
and','avg_goal_dif'])

df.plot(figsize=(12,5),title='Average Goals Difference Home vs Out')
```

Out[12]:

```
<AxesSubplot:title={'center':'Average Goals Difference Home vs Out'}>
```



In [ ]: